Reliable pre-eclampsia pathways based on multiple independent microarray data sets.

AUTHOR(S):

Kawasaki, Kaoru; Kondoh, Eiji; Chigusa, Yoshitsugu; Ujita, Mari; Murakami, Ryusuke; Mogami, Haruta; Brown, J. B.; Okuno, Yasushi; Konishi, Ikuo

CITATION:

Kawasaki, Kaoru ...[et al]. Reliable pre-eclampsia pathways based on multiple independent microarray data sets.. Molecular human reproduction 2014, 21(2): 217-224

ISSUE DATE:

2014-10-16

URL:

http://hdl.handle.net/2433/200675

Original

March 15, 2014

Reliable preeclampsia pathways based on multiple independent microarray data sets

Kaoru Kawasaki [1], Eiji Kondoh [1*], Yoshitsugu Chigusa [1], Mari Ujita [1],
Ryusuke Murakami [1], Haruta Mogami [1], J.B. Brown [2], Yasushi Okuno [2], Ikuo Konishi [1]

1 Department of Gynecology and Obstetrics, Kyoto University, Kyoto, Japan

2 Department of Clinical System Onco-Informatics, Kyoto University, Kyoto, Japan

Short title: Preeclampsia pathways and microarray analysis

* Corresponding Author
Department of Gynaecology and Obstetrics, Kyoto University
54 Shogoin Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan
Tel.: +81-75-751-3269, Fax: +81-75-761-3967
E- mail: kondo@kuhp.kyoto-u.ac.jp

Abstract

Preeclampsia is a multifactorial disorder characterized by heterogeneous clinical manifestations. Gene expression profiling of preeclamptic placenta have provided different and even opposite results, partly due to data compromised by various experimental artefacts. Here we aimed to identify reliable preeclampsia-specific pathways using multiple independent microarray data sets. Gene expression data of control and preeclamptic placentas were obtained from Gene Expression Omnibus. Single-sample gene-set enrichment analysis was performed to generate gene-set activation scores of 9,707 pathways obtained from the Molecular Signatures Database. Candidate pathways were identified by t-test-based screening using data sets, GSE10588, GSE14722, and GSE25906. Additionally, Recursive Feature Elimination was applied to arrive at a further reduced set of pathways. To assess the validity of the preeclampsia pathways, a statistically-validated protocol was executed using five data sets including two independent other validation data sets, GSE30186, GSE44711. Quantitative real-time PCR was performed for genes in a panel of potential preeclampsia pathways using placentas of 20 women with normal or severe preeclamptic singleton pregnancies (n=10, respectively). A panel of ten pathways were found to discriminate women with preeclampsia from controls with high accuracy. Among these

were pathways not previously associated with preeclampsia, such as the GABA receptor pathway, as well as pathways that have already been linked to preeclampsia, such as the glutathione and CDKN1C pathways. The mRNA expression of GABRA3 (GABA receptor pathway), GCLC and GCLM (glutathione metabolic pathway), and CDKN1C were significantly reduced in the preeclamptic placentas. In conclusion, ten accurate and reliable preeclampsia pathways were identified based on multiple independent microarray data sets. A pathway-based classification may be a worthwhile approach to elucidate the pathogenesis of preeclampsia.

Keywords: microarray; pathway; preeclampsia.

Introduction

Preeclampsia is a major cause of maternal and neonatal mortality and morbidity (Young BC1 et al., 2010). Preeclampsia is a heterogeneous syndrome in which the pathogenesis can be diverse among women (Young BC1 et al., 2010). Although the primary role of the placenta in the pathogenesis of preeclampsia is undisputed, its precise mechanism has yet to be fully elucidated. Consequently, the only definitive treatment for preeclampsia is delivery of the placenta, and no other effective therapy has been developed despite decades of extensive clinical and basic research. Thus, clearly there is an urgent need for clarification of the pathogenesis of preeclampsia.

Gene expression microarray data is a form of high-throughput genomics data for thousands of genes in each sample. Microarray-based gene expression profiling has provided numerous genes and pathways involved in preeclampsia (Sitras V et al., 2009: Winn VD et al., 2009: Tsai S et al., 2011: Louwen F et al., 2012). For example, Maynard et al. conducted gene expression profiling of placental tissue from women with and without preeclampsia, and found soluble fms-like tyrosine kinase 1 (sFlt1) (Maynard SE et al., 2003) to be closely related to the pathogenesis of preeclampsia. In addition, angiogenesis and immune-response pathways have been shown to be involved in preeclampsia in most microarray data sets (Sitras V et al., 2009: Winn VD et al., 2009: Tsai S et al., 2011: Louwen F et al., 2012). However, the genes and pathways derived from microarray analyses are diverse and even occasionally conflicting in existing studies (Winn VD et al., 2009: Tsai S et al., 2011: Louwen F et al., 2012). This might be attributed to sample differences in gestational age, modes of delivery, or experimental artefacts

such as types of chips and platform, as well as heterogeneous aetiologies or clinical manifestations. Thus, a single microarray data set may be insufficient to provide meaningful genes and pathways specific to preeclampsia. Indeed, more robust sets of genes and pathways have been provided through multiple independent data sets in a wide range of fields such as cancer research (Sorlie T et al., 2003: Rhodes DR et al., 2004). In the last decade, thousands of microarray data sets have appeared in public databases, which allow other researchers to confirm the results of published papers or to permit novel analyses of the data. Nevertheless, few studies (Moslehi R et al., 2013) have been conducted with the use of multiple data sets to seek genes and pathways in preeclamptic placentas. We hypothesized that pathways identified based on multiple independent microarray data sets from studies with large sample sizes were more likely to be functionally relevant to the pathogenesis of preeclampsia, and could potentially be new therapeutic targets for preeclampsia. The aim of our study was to provide preeclampsia-specific pathways using the three largest microarray data sets from four different platforms freely available in a web database.

Materials and Methods

Identification of common pathways overlapping three independent data sets in silico

In order to identify potentially relevant pathways to preeclampsia, gene expression data of control and preeclamptic placentas were obtained from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/gds/) as series matrix files. The selection criteria for the datasets were the datasets from the three largest sample sizes (sample size: GSE10588, 43; GSE14722,

23; GSE25906, 60) available at GEO DataSets, because the larger sample sizes can yield more reliable results. A summary of the analysed microarray data sets is shown in Table 1 (Sitras V et al., 2009: Winn VD et al., 2009: Tsai S et al., 2011: Meng T et al., 2012 : Blair JD et al., 2013). In GSE14722 study, the same samples were assayed on two different versions of the Affymetrix U133 arrays. The HG-U133A Array includes representation of the RefSeq database sequences and probe sets related to sequences previously represented on the Human Genome U95Av2 Array. In contrast, the HG-U133B Array contains primarily probe sets representing expressed sequence tag clusters. Thus, both of the two different versions of the Affymetrix U133 arrays are meant to be complementary and non-overlapping. HG-U133A and HG-U133B data were therefore combined for further analysis. Different types of microarray platforms have shown significant variability when comparing across platforms. Therefore, the three largest data sets from four different platforms were used for subsequent analysis. Single-sample gene-set enrichment analysis (ssGSEA) was performed to generate gene-set activation scores (Barbie DA et al., 2009). The ssGSEA script was obtained from GenePattern (http://www.broadinstitute.org/cancer/software/genepattern). According to the instructions described in ssGSEAProjection Documentation, v4 (http://www.broadinstitute.org/cancer/software/genepattern/modules/docs/ssGSEAProjection/4), GCT files containing the gene expression data were created as input files. Gene sets (8,513 pathways) were downloaded from the Molecular Signatures Database v3.1 (http://www.broadinstitute.org/gsea/downloads.jsp), and a “msigdb.v3.1.symbols.gmt” file, that consisted of all gene set collections named c1, c2, c3, c4, c5 and c6, was used for ssGSEA. We

added sets that combined up- and down-regulated sets derived from the same experimental condition or publication (option provided by ssGSEA package). The final total was therefore 9,707 pathways. Pathway activation scores in each sample were calculated using R (64 bit, 2.15.1) software (http://www.r-project.org/) using the “ssGSEAProjection.Library.R” and “common.R” scripts of the ssGSEA package, as shown in Supplementary Table 1. The t-test was used to compare the pathway activation scores between preeclampsia and control groups. Candidate preeclampsia pathways were explored using t-test-based screening of the 9,707 pathways on three independent microarray data sets (GSE10588, GSE14722, and GSE25906). In each data set, 180 pathways with top-ranked pathway activation scores were uniformly selected. Concordant candidate pathways in at least two data sets were considered as preeclampsia-specific pathways.

Validation of potential candidate pathways for preeclampsia in silico

In order to assess the resulting set of pathways on the ability to distinguish between preeclamptic and control cases, we executed 100 trials of Support Vector Machine (SVM) modeling and prediction, randomly splitting the samples into equal amounts of training and test data, for both endpoints for each trial (B. Schçlkopf et al., 2002). Hence, for example, in the case of experiments using the GSE10588, GSE14722, and GSE25906 datasets, 37 of the 74 control cases were randomly selected for use as training data, and the remaining 37 were held out for a prediction test; the 52 preeclampsia cases were handled similarly, and thus a training set contained 63 example cases along with a test set of 63 cases not included in the training data.

In each modeling trial, a SVM model was constructed after an automated parameter grid search using 3-fold cross-validation. This model was then used to predict the preeclampsia or control status of each case in the test data, and the model was evaluated using the accuracy [(TP+TN) / (TP+FP+TN+FN)], Area Under the ROC Curve (AUC) on the test data, and Matthews Correlation Coefficient [(TP*TN) - (FP*FN) / sqrt ((TP+FP) * (TP+FN) * (TN+FP) * (TN+FN))] (MCC) metrics.

In order to handle the per-batch effects of microarrays and resulting ssGSEA scores, two normalization procedures were executed for evaluation of modeling and analysis of results. In both cases, the normalization was done with respect to each pathway (ssGSEA score) using all samples in the batch processed. The first normalization procedure was to scale by using the sample mean and standard deviation, which is also known as the Z-scale transformation [(x – u) /s, u = sample mean, s = sample standard deviation]. The second normalization procedure was to apply an affine scaling by using the original range of values and scaling to the range [-1,1]. It is well known that the SVM algorithm performs better in general when data is scaled, so these two pathway score transformations are appropriate to the data and algorithm used in the study.

In total, four variations of randomized analysis on the reduced pathway set were executed. The reason for this is because we evaluated the statistical performance of modeling using the GSE10588, GSE14722, and GSE25906 datasets as well as when including the GSE30186 and GSE44711 studies (total of 5 datasets). The two types of combined datasets have nearly identical ratios of preeclampsia to control cases. For each type of combined dataset, the two aforementioned normalizations were applied before evaluation.

Systematic identification of a reduced set of critical pathways

Next, we executed a further analysis to assess if the focused set of pathways could be further reduced to an even smaller pathway subset which maintains predictive ability. For this purpose, we executed Recursive Feature Elimination (RFE), with a linear kernel Support Vector Machine as the modeling algorithm and feature (pathway) weighting mechanism (Isabelle Guyon et al., 2002). In RFE, the sample features are assigned weights during the model construction process, and features with lower weight are eliminated; this process is recursively done until the original number of features is reduced to a specified number of features. In this work, we eliminated one pathway per RFE pass.

As in the case with the randomized sample modeling, we executed RFE for four variations of datasets. The number of pathways was reduced by RFE to 10 for each dataset. The remaining pathways in each variation were tabulated and considered for their involvement in preeclampsia. Further, randomized sample modeling based on the RFE-reduced set of pathways was executed using the same protocol described above.

As an additional method of examining the results of RFE, we applied multi-dimensional scaling (MDS) to the further reduced datasets (J.B.Kruskal et al ., 1964). In short, MDS automatically derives coordinates for a series of datapoints, given a matrix of distances between each pair of datapoints. For visualization purposes, we calculated a MDS solution in two-dimensional space after transforming the post-RFE matrices to per-patient distances quantified by the standard Euclidean distance metric.

Patients and placenta samples

Twenty women with normal and severe preeclamptic singleton pregnancies were analysed in this study (n=10, respectively; Table 2). Severe preeclampsia was defined as maternal systolic blood pressure ≥ 160 mmHg and/or diastolic blood pressure ≥ 110 mm Hg in two consecutive measurements at least six hours apart, and proteinuria ≥ 2 g/24 h after 20 weeks of gestation. Small for gestational age was defined as relative birth weight less than the 10th percentile according to Japanese standards. Women with pre-existing chronic hypertension, renal disease, lupus erythematosus, diabetes or gestational hypertension without proteinuria were excluded.

Placental villous tissues were obtained immediately after Caesarean section in the absence of labour at Kyoto University Hospital, Japan. Villous tissues were collected from the central part of the placenta, and were macroscopically free of infarction or calcification. After brief rinsing in saline, these tissues were stored in RNAlater (Ambion, Austin, Texas) at -80ºC until RNA extraction. The study protocol was approved by the Ethics Committee, Graduate School and Faculty of Medicine, Kyoto University, and written informed consent was obtained from each patient.

Quantitative real-time PCR

Total RNA extraction from placental tissues was performed using an RNeasy Mini kit (QIAGEN, Germantown, Maryland). The quality and quantity of RNA was measured using an

ND-1000 spectrophotometer (Nanodrop, Wilmington, North Carolina). Reverse transcription of 1 mg RNA was performed using the Rever Tra Ace (TOYOBO, Osaka, Japan). The forward and reverse primers used for cDNA amplification are shown in Supplementary Table 2. Quantitative real-time PCR was performed using SYBR premix ExTaqII (Takara Bio, Otsu, Japan) on the LightCycler 480 Real-Time PCR system (Roche Diagnostics, Mannheim, Germany) as previously described (Chigusa Y et al., 2013).

Results

Pathway analysis based on independent data sets to discover preeclampsia-specific pathways

The results of comprehensive analysis of 9,707 pathways using t-test-based screening are available at XXXX. Of the top 180 pathways in each data set, only 21 pathways were common to at least two data sets (Supplementary Table 3). The panel of candidate pathways included well-known pathways involved in preeclampsia such as glutathione (oxidative stress), NF-kB (inflammation) and CDKN1C pathways. Moreover, the current study exhibited the emergence of novel pathways (e.g. GABA receptor and Sonic hedgehog) and potential susceptibility loci (3q and 4p15) for preeclampsia that have not been reported as being associated with preeclampsia. All of the genes involved in preeclampsia-specific pathways are shown in Supplementary Table 4.

Validation of potential candidate pathways for preeclampsia in silico

The results of randomized sampling and modeling using the reduced set of 21 pathways are as follows. The smaller combined dataset (GSE10588, GSE14722, and GSE25906) had an average accuracy of 84.6% +/- 4.9%, AUC-test of 0.980 +/- 0.014, and MCC of 0.691 +/- 0.092 using affine scaling; using Z-scaling, they had an average accuracy of 83.2% +/-4.3%, AUC-test of 0.975 +/- 0.016, and MCC of 0.664 +/- 0.085. The larger combined dataset (GSE10588, GSE14722, GSE25906, GSE30186, and GSE44711) had an average accuracy of 79.9% +/- 4.6%, AUC-test of 0.964 +/- 0.022, and MCC of 0.593 +/- 0.092 using affine scaling; applying the Z-scale transformation led to average accuracy of 80.8% +/- 4.5%, AUC-test of 0.965 +/- 0.018, and MCC of 0.616 +/- 0.089. From these results, we conclude that either type of normalization provides highly reasonable prediction performance, and the difference in prediction performance metrics as a function of dataset size is not dramatically altered. It suggests that the focused set of pathways related to prediction of preeclampsia is appropriate.

Randomized sample modeling based on the RFE-reduced set of 10 pathways (Table 3) was executed and evaluated. The results using the smaller combined set with affine scaling had an average accuracy of 86.2% +/- 4.4%, AUC-test of 0.982 +/- 0.014, and MCC of 0.720 +/- 0.090; the smaller dataset with Z-scaling resulted in an average accuracy of 83.6% +/- 4.8%, AUC-test of 0.976 +/- 0.013, and MCC of 0.669 +/- 0.085. The larger dataset normalized by affine scaling had an average accuracy of 82.0% +/- 3.5%, AUC-test of 0.971 +/- 0.015, and MCC of 0.636 +/- 0.072; Z-scale normalization yielded an average accuracy of 83.1% +/- 3.7%, AUC-test of 0.973 +/- 0.014, and MCC of 0.660 +/- 0.073. From such results and comparison

to the original random sampling experiment using 21 pathways, we observe that the 10 pathways remaining after RFE-SVM analysis continue to have a high discriminative ability for preeclampsia.

In Figure 1, the results of RFE-SVM analysis using the Z-scale transformation on the smaller combined dataset are shown. It is evident from the figure that the reduced set of 10 pathways is discriminative for preeclampsia, and motivates further study on the individual pathways and their involvement. A further analysis of only the preeclampsia patients in which they are clustered using the cosine distance with complete linkage is given as Supplementary Figure 1. Additionally, the original set of 9707 pathways visualized by means of MDS could not result in clearly distinguishable patient groups, but MDS visual analysis was much more successful with the reduced set of 10 pathways (see Supplementary Figure 2).

Quantitative real-time PCR for genes in preeclampsia-specific pathways

To validate the results obtained from pathway analysis, the expressions of selected genes involved in the glutathione metabolic pathway (GCLC and GCLM), CDKN1C pathway (CDKN1C), and GABA receptor pathway (GABRA3) were analysed by quantitative real-time PCR, respectively. The expression of each of these genes was significantly reduced in the preeclamptic placentas compared to controls (Figure 2), and these findings reinforce the data of pathway analysis based on independent data sets.

Discussion

Preeclampsia has diverse clinical manifestations such as mild or severe preeclampsia, early or late onset, and presence or absence of foetal growth restriction. Although previous studies using microarray analysis sought to find differentially expressed genes and pathways in preeclampsia, their results have been inconsistent (Sitras V et al., 2009: Winn VD et al., 2009: Tsai S et al., 2011: Louwen F et al., 2012: Meng T et al., 2012: Blair JD et al., 2013). This may be partly due to small numbers of study participants or differences in the microarray platform. In the current study, we used the independent data sets with the three largest sample sizes from four different platforms available as GEO datasets in order to avoid various biases. Initially, we tried to screen candidate pathways using false discovery rate (FDR). FDR is designed to prevent a large proportion of false positives, and is commonly used in the analysis of a large number of distinct variables in multiple samples. In the current case, there was only a single pathway left (KORKOLA_CHORIOCARCINOMA) common to at least two data sets (FDR<0.25). Thus, we did not use FDR as a method for screening for candidate preeclampsia pathways. Instead, we performed t-test-based screening.

We found that t-test-based screening under the following conditions (180 top-ranked pathways in GSE10588, GSE14722, and GSE25906) of the 9,707 pathways yielded only two pathways (IL2_UP.V1 and KORKOLA_CHORIOCARCINOMA) common to all of the three independent microarray data sets, suggestive of the heterogeneous genomic expression in preeclampsic placentas. Nevertheless, the pathway analysis also revealed that a panel of 21 identified pathways, as well as 10 pathways that were narrowed down using computational analysis, discriminates preeclamptic placentas from controls in not only the smaller combined

three data sets used to identify the pathways but also in larger data sets including two independent data sets despite various gestational ages, mode of delivery, and presence or absence of labour onset, indicating that these pathways are highly specific to the pathogenesis of preeclampsia.

To date, a single study alone has been reported with the use of multiple datasets from multiple data sources to seek genes and pathways involved in preeclampsia, but their study demonstrated computational analysis alone without sufficient validation (Moslehi R et al., 2013). Our study seems to have a number of strengths despite the conceptual similarity between their study and ours. First, our analysis included the largest placenta microarray study in preeclampsia (GSE25906). Second, we conducted pathway-based screening using a collection of pre-specified gene sets. Organizing genes into gene sets provides a more intuitive and stable context for assessing deeper biological insights in preeclampsia, because gene function is collectively exerted and may vary by environmental stimuli, or disease state. Finally, in order to confirm screening results, we conducted multiple validation through 100 trials of SVM modeling and prediction for both the smaller collection of three GSE datasets and the slightly larger collection of five GSE datasets. We found that results were quite similar regardless of either the collection and/or the method used to normalize data when compensating for per-batch effects. Additionally, we applied RFE to arrive at a further reduced set of pathways that contribute to the discriminative ability of a SVM to distinguish PE from control cases. Via RFE, we selected 10 pathways, and repeated the 100-trial random sampling and evaluation procedure. We found that performance was similar to the initial 100-trial experiment executed, signalling the importance of

the pathways selected by RFE. Furthermore, we performed confirmatory quantitative real time PCR for several selective genes related to candidate pathways using preeclamptic placentas and controls from our own institution. In reality, the results of microarray analysis are quite often unable to be verified in other datasets. Nevertheless, cluster analysis demonstrated that not only the initially reduced dataset (21 pathways) but also a further reduced dataset (10 pathways) discriminated preeclamptic placentas from controls irrespective of the smaller or larger dataset, and irrespective of the pathway score normalization procedure. Taken together, we believe that the panel of 10 pathways can provide deep biological insights into preeclampsia because our findings were based on multiple independent microarray data sets and deliberate validation.

Potential candidate functions or pathways that have been reported previously include angiogenesis, immune, inflammatory, oxidative stress, cell proliferation and differentiation, and metabolism (Sitras V et al., 2009: Winn VD et al., 2009: Tsai S et al., 2011: Louwen F et al., 2012: Meng T et al., 2012: Blair JD et al., 2013). Consistently, we identified ten preeclampsia-specific pathways which contained previously described pathways such as glutathione (Mistry HD et al., 2010), IL2 (Hmai et al., 1997) and CDKN1C (Kanayama N et al., 2002) pathways. Furthermore, we also discovered several novel pathways potentially involved in the pathogenesis of preeclampsia, such as GABA receptor and Sonic hedgehog pathways. After executing RFE analysis described above, we found that the GABA receptor, Sonic hedgehog, and 4p15 pathway were always selected as a relevant pathway. Hence, these newly identified pathways warrant further investigation.

Glutathione, a predominant intracellular antioxidant, is synthesized in the cytosol in a

tightly regulated manner. Mistry et al. reported that antioxidant enzyme glutathione peroxidase (GPx) is reduced in preeclamptic placentae (Mistry HD et al., 2010). In addition, we first found that GCLC and GCLM, both of which are rate-limiting enzymes in the biosynthesis of glutathione, were significantly decreased in preeclamptic placentae. Consistent with this, we previously reported that the activation of Nrf2, a predominant transcriptional factor of both GCLC and GCLM, was reduced in preeclamptic placentae (Chigusa Y et al., 2012). Furthermore, this is the first report that GABRA3 are suppressed in preeclamptic placentae. GABA receptors are associated with oxidative stress-induced apoptosis (Berntsent HF et al., 2013), and the activation of GABA receptor signalling reduces oxidative stress-mediated damage in liver (Gardner LB et al., 2012). These findings support the evidence that an impaired antioxidant defence system in the placenta is related to the pathogenesis of preeclampsia.

Preeclampsia is a multifactorial systemic vascular disorder affecting 5%–8% of all pregnancies. It has been suggested that immunologic factors cause failure of the trophoblast to sufficiently invade and remodel maternal uterine arteries at the fetomaternal interface (Redman CW et al., 2005), and that some are linked to a multifactorial polygenic inheritance with a genetic component (Redman CW et al., 2005: Arngrímsson R et al., 1999: Lachmeijer AM et al., 2001). A familial predisposition to preeclampsia has been demonstrated through previous studies which identified susceptibility loci for preeclampsia on 2p, 4q, 9p, 10q, 11q and 22q (Arngrímsson R et al., 1999: Lachmeijer AM et al., 2001: Laivuori H et al., 2003). In the present study, the loci on chromosome 3q and 4p15 were newly identified as candidate loci for preeclampsia.

Preeclamptic placenta and cancer share a number of common pathways including angiogenesis, immune, inflammatory, oxidative stress, cell proliferation and differentiation, and metabolic pathways (Louwen F et al., 2012). Although most cancer is quite heterogeneous in clinical phenotype as well as pathological findings, a pathway-based classification discovered subtypes that reflect specific histological properties and clinical outcomes in breast and lung cancer (Gatza ML et al., 2003: Nevins JR et al., 2011). We anticipate that this is also the case with preeclampsia. In the current study, some of the 10 pathways showed seemingly opposite directions and four subtypes may exist in preeclamptic cases (Supplementary Figure 1). For example, the heatmap of normalized pathway activation scores demonstrated that the Sonic hedgehog pathway or the glutathione pathway was down-regulated in most, but not all, samples from preeclamptic placentas. These findings are probably due to the heterogeneity of preeclampsia, and suggest that the pathway-based classification is likely to be a worthwhile approach to elucidate the pathogenesis of preeclampsia, and that preeclampsia could be categorized into clinically meaningful subtypes, including early/late onset, mild/severe preeclampsia, presence/absence of severe proteinuria, and coincident or not with foetal growth restriction, based on multiple distinct pathways. If detailed vital information could be obtained in each data set analysed in the current study, subpopulations of patients with common clinical manifestations might be identified using the panel of 10 pathways. The present study may be valuable in the understanding of the heterogeneity of preeclampsia and for providing a framework to develop rational therapeutic strategies according to pathway-based subtypes. On the other hand, the major limitation of the study is that this is basically an in-silico study using a

limited number of data sets including different modes of delivery, and presence or absence of labour onset.

In conclusion, ten accurate and reliable preeclampsia pathways were identified based on multiple independent microarray data sets.

Author Contributions

Kondoh E designed this study. Kawasaki K, Kondoh E, Murakami R, Brown J.B, and Okuno Y analyzed and interpreted data. Kawasaki K, Ujita M, Chigusa Y and Mogami H collected and assembled data. Konishi I finally approved the version to be published.

Disclosure

The authors report no conflict of interest.

Funding

This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Culture and Sports, Japan (No. 21592096, 22591822 and 23791833) and by a grant from the Smoking Research Foundation.

References

Arngrímsson R, Sigurõardóttir S, Frigge ML, Bjarnadóttir RI, Jónsson T, Stefánsson H, Baldursdóttir A, Einarsdóttir AS, Palsson B, Snorradóttir S, et al (1999) A genome-wide scan reveals a maternal susceptibility locus for pre-eclampsia on chromosome 2p13. Hum Mol Genet 8, 1799-805.

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, et al (2009) Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 5;462, 108-12.

Berntsen HF, Wigestrand MB, Bogen IL, Fonnum F, Walaas SI, Moldes-Anaya A (2013) Mechanisms of penitrem-induced cerebellar granule neuron death in vitro: possible involvement of GABAA receptors and oxidative processes. Neurotoxicology 35, 129-36.

Blair JD, Yuen RK, Lim BK, McFadden DE, von Dadelszen P, Robinson WP (2013) Widespread

DNA hypomethylation at gene enhancer regions in placentas associated with early-onset pre-eclampsia. Mol Hum Reprod 19, 697-708.

B. Schçlkopf, A. Smola (2002) Learning with Kernels: Support Vector Machines Regularization, Optimization, and Beyond. MIT Press, Cambridge.

Chigusa Y, Tatsumi K, Kondoh E, Fujita K, Nishimura F, Mogami H, Konishi I (2012) Decreased lectin-like oxidized LDL receptor 1 (LOX-1) and low Nrf2 activation in placenta are involved in preeclampsia. J Clin Endocrinol Metab 97:E, 1862-70.

Chigusa Y, Kondoh E, Mogami H, Nishimura F, Ujita M, Kawasaki K, Fujita K, Tatsumi K, Konishi I (2013) ATP-binding cassette transporter A1 expression is decreased in preeclamptic placentas. Reprod Sci 20, 891-8.

Gardner LB, Hori T, Chen F, Baine AM, Hata T, Uemoto S, Nguyen JH (2012) Effect of specific activation of γ-aminobutyric acid receptor in vivo on oxidative stress-induced damage after extended hepatectomy. Hepatol Res 42, 1131-40.

Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, Crawford MD, Datto MB, Kelley M, Mathey-Prevot B, Potti A, et al (2010) A pathway-based classification of human breast cancer. Proc Natl Acad Sci U S A 13;107, 6994-9.

Hamai Y, Fujii T, Yamashita T, Nishina H, Kozuma S, Mikami Y, Taketani Y (1997) Evidence for an elevation in serum interleukin-2 and tumor necrosis factor-alpha levels before the clinical manifestations of preeclampsia. Am J Reprod Immunol. Aug 38(2), 89-93.

Isabelle Guyon, Jason Weston, Stephen Barnhill, Vladimir Vapnik (2002) Gene Selection for Cancer Classification using Support Vector Machines Machine Learning 46:1-3, pp 389-422.

J.B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric Hypothesis (1964) Psychometrika 29, 1

J. Shawe-Taylor, N. Cristianini, Kernel (2004) Methods for Pattern Analysis, Cambridge University Press, Cambridge, UK.

Kanayama N, Takahashi K, Matsuura T, Sugimura M, Kobayashi T, Moniwa N, Tomita M, Nakayama K (2002) Deficiency in p57Kip2 expression induces preeclampsia-like symptoms in mice. Mol Hum Reprod 12, 1129-35.

Lachmeijer AM, Arngrímsson R, Bastiaans EJ, Frigge ML, Pals G, Sigurdardóttir S, Stéfansson H, Pálsson B, Nicolae D, Kong A, et al (2001) A genome-wide scan for preeclampsia in the Netherlands. Eur J Hum Genet 9, 758-64.

Laivuori H, Lahermo P, Ollikainen V, Widen E, Häivä-Mällinen L, Sundström H, Laitinen T, Kaaja R, Ylikorkala O, Kere J (2003) Susceptibility loci for preeclampsia on chromosomes 2p25 and 9p13 in Finnish families. Am J Hum Genet 72, 168-77.

Louwen F, Muschol-Steinmetz C, Reinhard J, Reitter A, Yuan J (2012) A lesson for cancer research: placental microarray gene analysis in preeclampsia. Oncotarget 3, 759-73.

Maynard SE, Min JY, Merchan J, Lim KH, Li J, Mondal S, Libermann TA, Morgan JP, Sellke FW, Stillman IE, et al (2003). Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia. J Clin Invest 111, 649-58.

Meng T, Chen H, Sun M, Wang H, Zhao G, Wang X (2012) Identification of differential gene expression profiles in placentas from preeclamptic pregnancies versus normal pregnancies by

DNA microarrays. OMICS, 16:301-11.

Mistry HD, Kurlak LO, Williams PJ, Ramsay MM, Symonds ME, Broughton Pipkin F (2010) Differential expression and distribution of placental glutathione peroxidases 1, 3 and 4 in normal and preeclamptic pregnancy. Placenta, 31:401-8.

Moslehi R, Mills JL, Signore C, Kumar A, Ambroggio X, Dzutsev A (2013) Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. Sci Rep. 3, 2407.

Nevins JR (2011) Pathway-based classification of lung cancer: a strategy to guide therapeutic selection. Proc Am Thorac Soc 8, 180-2.

Redman CW, Sargent IL (2005) Latest advances in understanding preeclampsia. Science 10;308, 1592-4.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci USA 22;101, 9309-14.

Sitras V, Paulssen RH, Grønaas H, Leirvik J, Hanssen TA, Vårtun A, Acharya G (2009) Differential placental gene expression in severe preeclampsia. Placenta 30, 424-33.

Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, et al (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100, 8418-23.

Tsai S, Hardison NE, James AH, Motsinger-Reif AA, Bischoff SR, Thames BH, Piedrahita JA

(2011) Transcriptional profiling of human placentas from pregnancies complicated by preeclampsia reveals disregulation of sialic acid acetylesterase and immune signalling pathways. Placenta, 32:175-82.

Young BC1, Levine RJ, Karumanchi SA (2010) Pathogenesis of preeclampsia. Annu Rev Pathol 5, 173-92.

Winn VD, Gormley M, Paquet AC, Kjaer-Sorensen K, Kramer A, Rumer KK, Haimov-Kochman R, Yeh RF, Overgaard MT, Varki A, et al (2009) Severe preeclampsia-related changes in gene expression at the maternal-fetal interface include sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2. Endocrinology 150, 52-62.

Figure legends

Figure 1. Cluster analysis using a panel of 10 preeclampsia pathways

Heatmap of normalized pathway activation scores using combined dataset (GSE10588, GSE14722, and GSE25906). The results of RFE-SVM analysis using the Z-scale transformation are shown. In the heatmap, each column represents one pathway, and each row corresponds to a sample of placenta. The relative score of each sample to the pathway is represented by a colour. High and low scores are shown in yellow and blue, respectively.

Figure 2. Expression of preeclampsia pathway-related genes in placenta

Validation of pathway analysis of microarray data by quantitative real-time PCR. Genes involved in glutathione metabolic pathway (GCLC and GCLM), CDKN1C pathway (CDKN1C), and GABA receptor pathway (GABRA3) were significantly down-regulated in the preeclamptic placentas compared to controls (n=10 in each group, Mann–Whitney U test). Data are shown as mean relative expression + SEM.

Supplementary Figure 1. Cluster analysis using a panel of 10 preeclampsia pathways in cases of preeclampsia

Heatmap of analysis of only the preeclampsia patients in which they are clustered using the cosine distance with complete linkage on ssGSEA scores normalized using the Z-scale transformation (GSE10588, GSE14722, and GSE25906).

Supplementary Figure 2. MDS visual analysis based on pathways

The reduced set of 10 pathways (upper panel) can discriminate control from preeclamptic placentas more clearly than the original set of 9707 pathways (lower panel).

Table 1. Summary of analysed microarray data sets

Dataset	Platform	Number of probes	Sample size	Gestaional age (wks)	labored	Cesarean delivery	Fetal gender	Year	Reference
GSE10588	ABI Human Genome Survey Microarray Version 2	32878	17 severe preeclampsia 26 control	34.0 ± 3.6 (n=16 †) 39.6 ± 1.3 (n=21 †)	5/16 † N/A	11/16 † 8/21†	N/A N/A	2009	2
GSE14722	Affymetrix Human Genome U133A Array Affymetrix Human Genome U133B Array	22115 22477	12 severe preeclampsia 11 control (preterm)	31.0 ± 4.6 32.1 ± 3.3	10/12 11/12	6/12 2/11	N/A N/A	2009	3
GSE25906	Illumina human-6 v2.0 expression beadchip	48701	23 preeclampsia 37 contol	34.2 ± 3.6 37.7 ± 2.0	16 induced/23 8 induced/37	N/A N/A	10 male, 13 female 21 male, 16 female	2010	4
GSE30186	Illumina HumanHT-12 V4.0 expression beadchip	47231	6 preeclampsia 6 control	36.4 ± 0.9 39.0 ± 0.7	0/6 0/6	6/6 6/6	N/A N/A	2012	9
GSE44711	Illumina HumanHT-12 V4.0 expression beadchip	47231	8 early-onse preeclampsia 8 control	32.2 ± 3.5 31.4 ± 3.9	N/A N/A	N/A N/A	6 male, 2 female 6 male, 2 female	2013	10

†, Data are shown as described in the article.

Table 3. Preeclampsia-specific pathways based on multiple independent microarray data sets

Pathway
KORKOLA_CHORIOCARCINOMA
BIOCARTA_SHH_PATHWAY
ISHIDA_TARGETS_OF_SYT_SSX_FUSIONS
chr4p15
REACTOME_GABA_RECEPTOR_ACTIVATION
GNF2_CDKN1C
KEGG_BUTANOATE_METABOLISM
IL2_UP.V1
CYCLASE_ACTIVITY
KEGG_GLUTATHIONE_METABOLISM

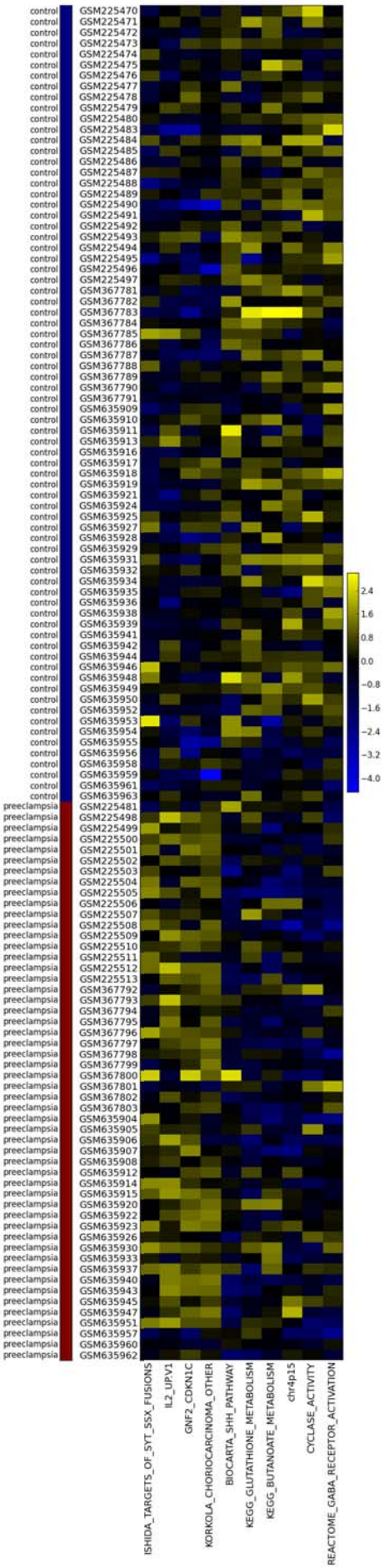
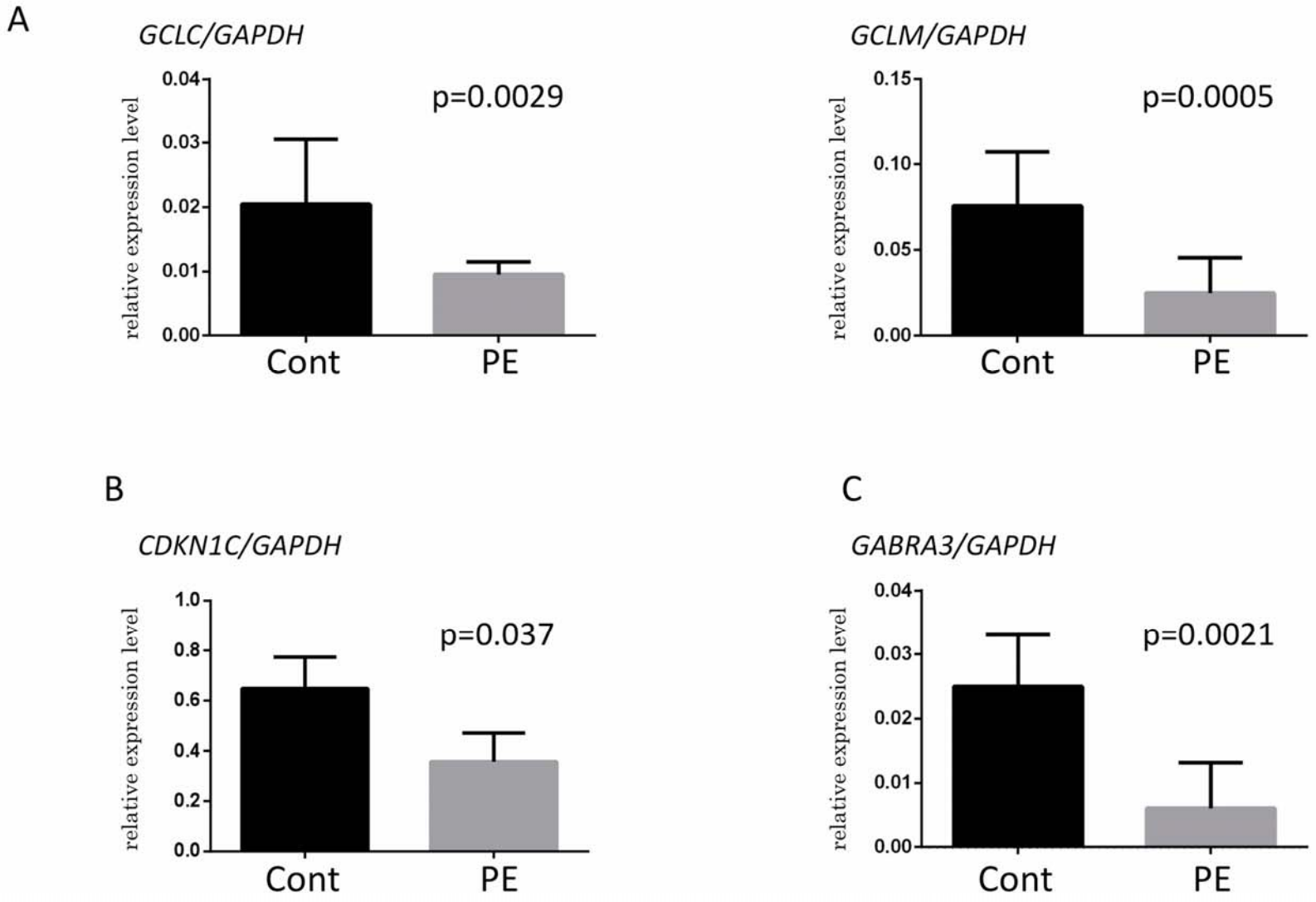


Figure 1

Figure 2


A
GCLC/GAPDH
relative expression level
0.04
0.03
0.02
0.01
0.00
p=0.0029
Cont
PE
GCLM/GAPDH
relative expression level
0.15
0.10
0.05
0.00
p=0.0005
Cont
PE
B
CDKN1C/GAPDH
relative expression level
1.0
0.8
0.6
0.4
0.2
0.0
p=0.037
Cont
PE
C
GABRA3/GAPDH
relative expression level
0.04
0.03
0.02
0.01
0.00
p=0.0021
Cont
PE

Supplementary Table 1

R script for ssGSEA analysis.

```
#ssGSEA R code
source("common.R")
source("ssGSEAProjection.Library.R")
ssGSEA<-ssGSEA.project.dataset(
	#javaexec,
	#jardir,
	input.ds= "Input_file_name.gct", # change the name
	output.ds= "Output_file_name.gct", # change the name
	gene.sets.database= "",
	gene.sets.dbfile.list= "msigdb.v3.1.symbols.gmt", # Geneset.gmt
	gene.symbol.column= "Description", # Description column contains gene symbol
names.
	gene.set.selection = "ALL",
	sample.norm.type	= "rank",
	weight			= 0.75,
	combine.mode		= "combine.add",
	min.overlap		= 1)
```

Supplementary Table 2

Primer sequences used in quantitative real-time PCR.

Gene	Forward	Reverse	Accession Number
GCLC	GGCACAAGGACGTTCTCAAGTG	CCATACTCTGGTCTCCAAAGGGTAG	NM_001498.2
GCLM	CCCAGATTTGGTCAGGGAGTTTCCA	ACTGAACAGGCCATGTCAACTGCA	NM_002061.2
CDKN1C	GGCCTCTGATCTCCGATTTCTTCG	GGGGCTCTTTGGGCTCTAAATTGG	NM_000076.2
GABRA3	TTTGGGCCATGTTGTTGGGACAGA	ACTCTCTGTTGAGCCAGAACGACAC	NM_000808
GAPDH	GAGTCAACGGATTTGGTCGTATTGG	GCCATGGGTGGAATCATATTGGAAC	NM_002046.3

Supplementary Table 6. Preeclampsia-specific pathways based on multiple independent microarray data sets

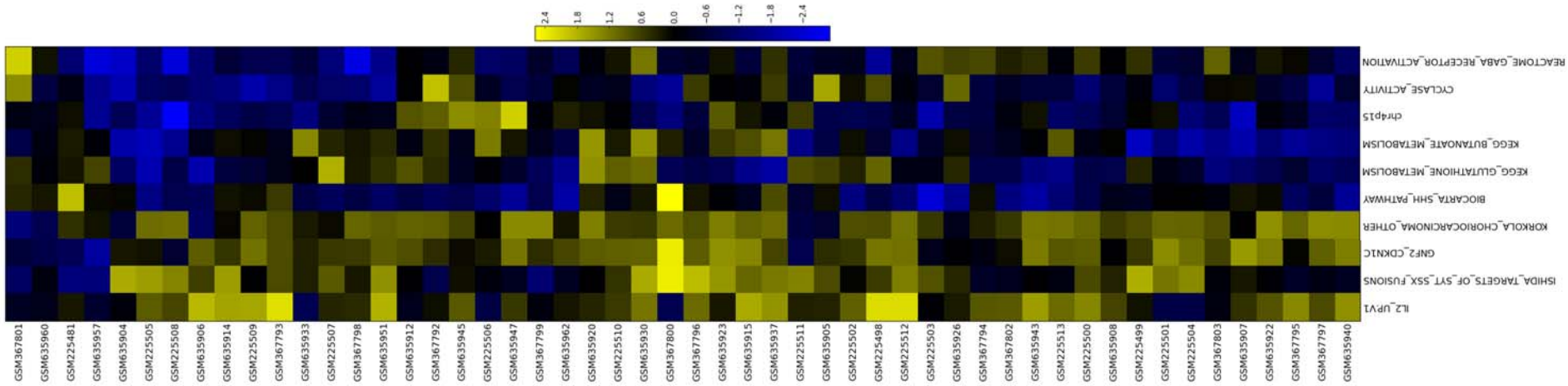
Pathway	GSE10588	GSE14722	GSE25906
KORKOLA_CHORIOCARCINOMA	**0.0001**	**0.0000**	**0.0033**
BIOCARTA_SHH_PATHWAY	**0.0001**	0.3820	**0.0014**
DACOSTA_UV_RESPONSE_VIA_ERCC3_XPCS	**0.0001**	0.5785	**0.0029**
ISHIDA_TARGETS_OF_SYT_SSX_FUSIONS	**0.0001**	0.8588	**0.0069**
MEISSNER_NPC_ICP_WITH_H3_UNMETHYLATED	**0.0002**	**0.0008**	0.7354
chr4p15	**0.0002**	**0.0097**	0.0424
RESPONSE_TO_NUTRIENT_LEVELS	**0.0002**	**0.0013**	0.1869
NEGATIVE_REGULATION_OF_RESPONSE_TO_STIMULUS	**0.0002**	**0.0009**	0.4361
LIM_MAMMARY_LUMINAL_PROGENITOR_DN	**0.0003**	**0.0106**	0.7399
REACTOME_GABA_RECEPTOR_ACTIVATION	**0.0005**	0.5389	**0.0075**
KEGG_BUTANOATE_METABOLISM	**0.0008**	**0.0049**	0.7086
IL2_UP.V1	**0.0008**	**0.0096**	**0.0000**
CYCLASE_ACTIVITY	**0.0008**	0.0541	**0.0040**
DACOSTA_UV_RESPONSE_VIA_ERCC3_XPCS_UP	**0.0009**	0.8480	**0.0016**
MANTOVANI_NFKB_TARGETS	**0.0009**	0.6597	**0.0065**
GNF2_CDKN1C	0.0110	**0.0045**	**0.0001**
MYELOID_CELL_DIFFERENTIATION	0.1252	**0.0029**	**0.0091**
chr3q	0.3021	**0.0073**	**0.0004**
NEGATIVE_REGULATION_OF_PHOSPHATE_METABOLIC_PROCESS	0.4665	**0.0001**	**0.0079**
KEGG_GLUTATHIONE_METABOLISM	0.6358	**0.0041**	**0.0050**
PID_P38ALPHABETAPATHWAY	0.9364	**0.0036**	**0.0088**

180 top-ranked pahways in each dataset are shown in bald.

Supplementary Table 7. Genes involved in 21 pathways.

Pathway																										
KORKOLA_CHORIOCARCINOMA	CGA	HSD3B1	**LEP**	LHB	**TFPI**	XAGE3																				
BIOCARTA_SHH_PATHWAY	DYRK1A	DYRK1B	GLI1	GLI2	GLI3	GSK3B	PRKACB	PRKACG	PRKAR1A	PRKAR1B	PRKAR2A	PRKAR2B	**PTCH1**	SHH	SMO	SUFU										
DACOSTA_UV_RESPONSE_VIA_ERCC3_XPCS	ABCC1	ABCE1	ACAP2	**ADCY9**	AGFG1	AHDC1	AKAP10	APBB2	ARAP2	ARHGEF10	ASXL1	ATP2C1	ATP8B1	ATRN	AVL9	BDNF	BICD1	BMPR1A	BTRC	CDH2	CENPC1	CTIF	CUL2	DKK1	DLEU2	DLG1
	DOCK4	DOCK9	DST	DUSP5	**E2F5**	EIF2C2	EIF3A	FAM155A	FAM168A	FAM179B	FAM193A	GPATCH8	GRK5	HEG1	HERC4	HOXB2	IGF1R	IL6	INTS3	KIAA0182	KIAA0922	MALT1	MPHOSPH9	MSH6	MTAP	MYO9B
	NAV3	NFATC3	NFKB1	PCCA	PCNT	PDLIM5	PDS5B	PHF14	PIK3C2A	PLCB4	PLCE1	PTEN	PTX3	PVRL3	RALGAPB	RANBP2	RB1CC1	RPS6KA3	SFMBT1	SLC25A12	SLIT2	SMAD4	SON	SOS2	SP100	TERF1
	TFPI	TGFBR3	TMCC1	TSC22D2	UBXN7	VLDLR	VPS13B	WDR37	WWP1	ZEB2																
ISHIDA_TARGETS_OF_SYT_SSX_FUSIONS	B4GALT1	BAX	DDIT3	GDF15	HYOU1	NUPR1																				
MEISSNER_NPC_ICP_WITH_H3_UNMETHYLATED	GPT	COL6A2	EPN3	EPPK1	EVPL	FOXS1	H1FNT	KRT85	LECT1	**LEP**	LYL1	MOGAT1	NCKAP5	PAK6	PALM3	PLEKHG4	PPIH	PROKR1	S1PR4	SHANK2	TNFRSF13C	TSSK3	TUBA3C	ZNF599		
chr4p15	HGD	ANAPC4	ARAP2	ATP1B1P1	ATP5LP3	BST1	C1QTNF7	C4orf52	CC2D2A	CCDC149	CCKAR	CD38	CLRN2	CPEB2	DCAF16	DHX15	DKFZp547J222	EPPS	FAM200B	FBXL5	FGFBP1	FLJ39653	FLJ45721	FRA4D	GBA3	GPR125
	GRXCR1	HLN2	HSP90AB2P	KCNIP4	KCNIP4-IT1	KLHL5	LAP3	LCORL	LGI2	LOC133185	LOC152742	LOC285540	LOC285547	LOC285548	LOC391636	LOC391640	LOC391642	LOC441009	LOC643446	LOC643751	LOC644753	LOC644816	LOC644868	LOC645108	LOC645433	LOC645481
	LOC645716	LOC727819	LOC727823	LOC729006	LOC729071	LOC729175	MAPRE1P2	MESTP3	MRPL51P1	NCAPG	OB4	PARK4	PCDH7	PI4K2B	PPARGC1A	PROM1	QDPR	RBPJ	RPL21P46	RPL31P31	RPS7P6	SEL1L3	SEPSECS	SLC34A2	SLIT2	STIM2
	TAPT1	TBC1D1	TBC1D19	UGDH	USP17	ZCCHC4																				
RESPONSE_TO_NUTRIENT_LEVELS	ALB	**ASNS**	CARTPT	CCKAR	**CDKN2B**	**CDKN2D**	CHMP1A	ENPP1	ENSA	FADS1	GCGR	**GHRL**	**GHSR**	GIPR	**GNAI2**	**LEP**	NPY	NUAK2	OGT	PCSK9	PPARG	SREBF1	SST	SSTR1	SSTR2	STC1
	STC2	TP53	TULP4																							
NEGATIVE_REGULATION_OF_RESPONSE_TO_STIMULUS	CARTPT	CHRNA7	FOXP3	**GHRL**	**GHSR**	KLK8	**LEP**	PTPRC	SPINK5	TARBP2	TGFB2															
LIM_MAMMARY_LUMINAL_PROGENITOR_DN	ASPH	CERS6	CUX1	DST	EPDR1	HSPB8	ITGA5	KCNIP3	KIF3C	PLCH2	SCARF2	SEMA3F	TBC1D9	TRIM29												
REACTOME_GABA_RECEPTOR_ACTIVATION	ADCY1	ADCY2	ADCY3	ADCY4	ADCY5	ADCY6	**ADCY7**	**ADCY8**	**ADCY9**	ARHGEF9	GABBR1	GABBR2	GABRA1	GABRA2	GABRA3	GABRA4	GABRA5	GABRA6	GABRB1	GABRB2	GABRB3	GABRG2	GABRG3	GABRR1	GABRR2	GNAI1
	GNAI2	GNAI3	GNAL	GNB1	GNB2	GNB3	GNG10	GNG12	GNG2	GNG3	GNG4	GNG5	GNG7	GNG8	GNGT1	GNGT2	KCNJ10	KCNJ12	KCNJ15	KCNJ16	KCNJ2	KCNJ3	KCNJ4	KCNJ5	KCNJ6	KCNJ9
KEGG_BUTANOATE_METABOLISM	AASC	ABAT	ACADS	ACAT1	ACAT2	ACSM1	ACSM2A	ACSM3	ACSM4	ACSM5	AKR1B10	ALDH1B1	ALDH2	ALDH3A2	ALDH5A1	ALDH7A1	ALDH9A1	BDH1	BDH2	ECHS1	EHHADH	GAD1	GAD2	HADH	HADHA	HMGCL
	HMGCS1	HMGCS2	L2HGDH	OXCT1	OXCT2	PDHA1	PDHA2	PDHB																		
IL2_UP.V1	AATK	ABTB2	ADAM19	AHR	AIM2	AK4P3	ALDH4A1	AMIGO2	ANXA2P1	AP3M2	ASCL3	ASPA	BMP2	BTC	C10orf2	C11orf80	C12orf44	C13orf15	C19orf28	CALML4	CAPN10	CCDC85B	CCR4	CD3EAP	CD52	CD69
	CDC25A	CDCP1	CDH3	CISH	CLDN5	CNKSR2	CXCL12	CXCL3	DGKG	DHRS3	DUSP2	DUSP4	DUSP6	EGR1	EGR2	EGR3	EHD2	ELMO3	ESRP2	EVPL	EVX1	F2RL2	FAM57A	FLJ11827	FLT1	FLT3LG
	FOS	FOSL1	FOSL2	FOXP3	FZD5	GABARAPL3	GABRB3	GADD45B	GALNT8	GFAP	GNRH2	**GPT**	HABP4	HEG1	HK2	HOXA11	HTR4	IFNW1	IL10	IL13	IL18RAP	IL1A	IL5	INHBB	INSR	IQCG
	IRS1	ITGA9	JUNB	KCNA3	KCNE1L	KCNG1	KCNK5	KCTD12	KIF13A	KIF1C	KIF21B	KIR2DL5A	KLF6	KLRK1	LGALSL	LIF	LILRB4	LMO3	LOC388796	LOC96610	LRP8	LRRC6	MAFF	MAGIX	MEGF6	MPP6
	NCS1	NR2F2	OR10J1	OR6A2	OR7E47P	OSM	PAK6	PDE4A	PIM1	PIM2	PIP	PLEC	PLXNA3	PMM2	POLR3G	PPAP2C	PPP4R4	PQLC2	PRAF2	PRKCD	PRR7	PRSS50	**PTCH1**	PTH1R	PTPN7	PUS1
	PUS7	PYCRL	RAB11FIP1	RAC3	REST	RGS16	RGS3	RHBDF2	RHOB	RNF144A	RNF20	RPA4	RPP25	RRP7B	RSPH6A	SH2D4A	SIGLEC9	SLAMF1	SLC1A5	SLC29A2	SLC2A3	SLC34A1	SLC7A5	SLCO4A1	SOCS1	SOCS2
	SPATA2L	SPHK1	SPP1	SPRY2	STX11	TAS2R16	TBC1D9	TEX13A	TLR3	TM4SF1	TM4SF5	TMEM158	TMPRSS6	TNFAIP8	TNFRSF12A	TNFRSF1B	TNFRSF21	TNFRSF4	TNFRSF9	TNFSF11	TNFSF14	TPBG	TRIB1	TRMU	TTTY1	URB2
	USH2A	USP36	VDR	VSNL1	WDR62	WISP3	YRDC	ZBTB32	ZNF215	ZNF343																
CYCLASE_ACTIVITY	PCYT1A	**ADCY7**	**ADCY8**	**ADCY9**	GUCY1A2	GUCY1A3	GUCY1B3	GUCY2C	GUCY2D	GUCY2F	RTCD1															
DACOSTA_UV_RESPONSE_VIA_ERCC3_XPCS_UP	ADO	ALDH1B1	CASP3	CCNT2	**CDKN1C**	CTH	FBXL14	FRAT2	GLS	GSTT2	HAUS3	IFI30	IRF1	LACRT	MED6	MPDU1	PCDH9	PDCL	RAD9A	RPUSD2	SAT1	SRSF9	TERF2	TMX4	TP53	TRIM21
	WBSCR22	ZKSCAN1																								
MANTOVANI_NFKB_TARGETS	**AP4B1**	BCL2	C1S	C3	CA2	CASP4	CCDC64	CCL2	CCL7	CDC14B	CP	CXCL1	CXCL12	CXCL9	DUSP6	FLT1	FSD1	GTF3C1	IGFBP7	IL10	INHBB	LATS2	MTRF1	MYC	NFKBIZ	OPN4
	OSMR	PLD4	PPP1R1B	PRKCE	SAA1	SLC16A4	SLC26A9	SLC29A1	SPP1	STEAP4	SYN3	TRIM2	TRPV6	TTYH3	VCAM1	VEGFA	ZCWPW1									
GNF2_CDKN1C	**ADAM12**	**ALPP**	**CAPN6**	**CDKN1C**	**CRH**	**CYP19A1**	**EGFL6**	**GCM1**	**GDF15**	**GH2**	**HSD17B1**	**HSD3B1**	**KISS1**	**LEP**	MAFF	MAN1C1	**PAPPA2**	**PSG1**	**PSG2**	**PSG3**	**PSG4**	**PSG5**	**PSG7**	**PAPPA3**	SEMA3B	**SVEP1**
	TIMP2																									
MYELOID_CELL_DIFFERENTIATION	ACIN1	ACVR1B	ACVR2A	ALAS2	CALCA	CARTPT	CDC42	CDK6	CEBPG	CSF1	DYRK3	ETS1	FOXO3	HCLS1	IFI16	IL31RA	IL4	**INHA**	**INHBA**	KAT6A	KAT8	LDB1	LYN	MAFB	MMP9	MYH9
	NCOA6	PF4	RASGRP4	RPS19	RUNX1	SCIN	SNRK	SPI1	TM7SF4	ZBTB16	ZNF675															
chr3q	**CBLB**	CLDN16	CLDN18	**HGD**	PCYT1A	SLC33A1																				
NEGATIVE_REGULATION_OF_PHOSPHATE_METABOLIC_PROCESS	CDKN1A	CDKN1B	CDKN2A	**CDKN2B**	CDKN2C	**CDKN2D**	IGFBP3	**INHA**	**INHBA**	NF2	NLRP12	PPP2R4	SOCS1													
KEGG_GLUTATHIONE_METABOLISM	ANPEP	G6PD	GCLC	GCLM	GGCT	GGT1	GGT5	GGT6	GGT7	GPX1	GPX2	GPX3	GPX4	GPX5	GPX6	GPX7	GSR	GSS	GSTA1	GSTA2	GSTA3	GSTA4	GSTA5	GSTK1	GSTM1	GSTM2
	GSTM3	GSTM4	GSTM5	GSTO1	GSTO2	GSTP1	GSTT1	GSTT2	GSTZ1	IDH1	IDH2	LAP3	MGST1	MGST2	MGST3	ODC1	OPLAH	PGD	RRM1	RRM2	RRM2B	SMS	SRM	TXNDC12		
PID_P38ALPHABETAPATHWAY	BLK	CCM2	CDC42	DUSP1	DUSP10	DUSP16	DUSP8	FGR	FTN	HCK	LCK	YKN	MAP2K3	MAP2K4	MAP2K6	MAP3k12	MAP3K3	MAPK11	MAPK14	PAK1	PAK2	PAK3	PPKG1	RAC1	RALA	RALB
	RIPK1	SRC	TAB1	TRAF6	YES1																					

Genes common to more than one pathway are shown in bald.



Supplementary Figure 1

Supplementary Figure 2